LCA Methodology

Comparative LCA of Industrial Objects

Part 1: LCA Data Quality Assurance – Sensitivity Analysis and Pedigree Matrix

Anna Lewandowska*, Zenon Foltynowicz and Andrzej Podlesny

Faculty of Commodity Science, The Poznan University of Economics, al. Niepodleglosci 10, 60-967 Poznan, Poland

* Corresponding author (lewandowska.anna@wp.pl)

Part 1: LCA Data Quality Assurance – Sensitivity Analysis and Pedigree Matrix
[Int J LCA 9 (2) 86 – 89 (2004)]
Part 2: Case Study for Chosen Industrial Pumps
[Int J LCA, OnlineFirst: DOI: http://dx.doi.org/10.1065/lca2004.03.152.2]

Preamble: This series consists of two papers which concern the same LCA case study but present two different problems. The first part is devoted to the quality assessment aspects and presents a proposal of the integration of sensitivity analysis and pedigree matrix in one analysis. The second part presents in detail the results of the comparative LCA case study for industrial pumps. The discussed theoretical considerations are implemented.

DOI: http://dx.doi.org/10.1065/lca2004.03.152.1

Abstract

Goal, Scope and Background. The main aim of this paper is to present some methodological considerations concerning existing methods used to assess quality of the LCA study. It relates mainly to the quality of data and the uncertainty of the LCA results. The first paper is strictly devoted to methodological aspects whereas, the second is presented in a separate article (Part II) and devoted mainly to a case study.

Methods. The presented analysis is based on two well-known concepts: the Data Quality Indicators (DQIs) and the Pedigree Matrix. In the first phase, the Sensitivity Indicators are created on the basis of the sensitivity analysis and then linked with the DQIs and the Quality Classes. These parameters indicate the relative importance of input data and their theoretical quality levels. Next, the Weidema's Pedigree Matrix (slightly modified) is used to establish the values of the new parameter called the Data Quality Distance (DQD) and to link them with the DQIs and Quality Classes. This way the information about the "real" quality levels is provided. Further analysis is performed using the probabilistic distributions and Monte Carlo simulations.

Results and Discussion. Thanks to this approach it is possible to make a comparison between two types of the quality factors. On the one hand, the sensitivity analysis allows one to check the importance of input data and to determine their required quality. It is done according to the following relation: the higher the sensitivity indicator, the higher the importance of input data and the higher quality should be demanded. On the other hand, the data have a certain real quality, not always in accord with the demanded one. To make possible a comparison between these two types of quality, it is necessary to find and develop a common denominator for them. Here, for this purpose the DQIs and Quality Classes are used.

Conclusions. In the further stage of the assessment the DQIs are used to perform the uncertainty analysis of the LCA results. The results could be additionally analysed by using other techniques of interpretation: the sensitivity-, the contribution-, the comparative-, the discernability- and the uncertainty analysis.

Recommendations and Outlook. The presented approach is put into practice to conduct the comparative LCA study for the industrial pumps by using the Ecoindicator99 method. Thanks to this, complex analysis of the credibility of the results is carried out. As a consequence, uncertainty ranges for the LCA results of every product system can be determined [1].

Keywords: Data quality distance (DQD); data quality goals (DQG); data quality indicators (DQI); pedigree matrix; quality classes; sensitivity analysis; uncertainty analysis

1 Introduction

The Life Cycle Assessment method has a lot of potential sources for the uncertainty [2]. Each of them influences the final results and the entire quality of the LCA study. To make a comparison between the results of the different LCA studies, it is necessary to make sure they are comparable not only in the modelling of the product systems (a definition of a function and a functional unit, the system boundaries, the value choices, the calculation procedure, a Life Cycle Impact Assessment methodology, etc.), but also in the levels of their uncertainty. One of the most important issues is the problem of data quality in LCA. It has been a field of a lively interest for many years, from the beginnings of working on LCA to present times. One of the first initiatives undertaken in this matter was the SETAC workshop in Wintergreen in 1992 [3]. The first mention about Data Quality Indicators (DQIs) and Data Quality Goals can be found in this report. The concept of DQIs has been improved in different ways: from the qualitative (U.S. EPA 1995) through the semi-quantitative (Weidema and Wesnoes 1995; Kennedy et al 1996) to the quantitative examples (SETAC 1994) [4–6]. In April 1998, the SETAC-Europe Working Group has launched a project called 'Data Availability and Data Quality'. Almost at the same time, work on the probabilistic approaches were presented by Heijungs (1996), Meier (1997) and Kennedy et al.(1997) [5]. The analysis presented in the article is mainly based on the results of Kennedy et al. (subsequently applied by Kusko and Hunt 1997) consisting of adopting a semi-quantitative approach to develop the uncertainty distributions for inventory data [6]. The data quality problem is inherently connected with the uncertainty and variability issues [8–10]. In this paper, the attention is focused on the above-mentioned concepts of DQIs and Weidema's Pedigree Matrix [4] in order to improve a probabilistic approach [9] and perform uncertainty analysis by transforming a deterministic model into a stochastic one [11]. Additionally, the sensitivity analysis is used to reach some minimal levels of data quality. One can find a lot of suggestions in [12–15] about how to use sensitivity analysis in the LCA process. The presented assessment has been implemented

into the comparative LCA analysis of pumps (which is the subject of a separate paper), that is why some comments and references to this study can be found in this work.

2 Sensitivity Analysis and Sensitivity Indicators

There is a rule that the quality of the LCA results will never be higher than the quality of the data used for a construction of Life Cycle Inventory and Life Cycle Impact Assessment models. One of the major weaknesses of the LCA is using data of very poor quality, sometimes even data points. No information about their uncertainty is given in such case. On the one hand, site-specific data differentiated according to time, technology and space can be used. On the other hand, the same data can come from an unrepresentative sample. The first question in this analysis is: what is the minimal quality level that could be accepted for each input data? The more important the data input, the higher the level of quality which should be required. For this reason, the level of data importance should be evaluated. That is why a sensitivity analysis is used. The input data are changed first by 1 percent and then by 10 percent and the reaction (change in percents) of the final results (the level of the environmental impact) is checked. Based on the degree of the reaction, the Sensitivity Indicators (SIs) are determined. Five types of SIs are distinguished:

- HS (High Sensitivity)
- LS (Low Sensitivity)
- LIS (Low Insensitivity)
- HIS (High Insensitivity)
- VHIS (Very High Insensitivity)

The fundamental issue is to make a link between SIs and the results of the sensitivity analysis. The analysis is carried out for the first-order input data and the change of the values of the Ecoindicator (in percent) is observed in all cases. The calculation is made in two ways: for the separate life cycle stages and for the whole life cycle. The obtained results are very interesting. There is a clear dominance of the results without any reaction (zero percent). It would mean that almost half of the data (about 48 percent) have a very small share in the final environmental impact, because even a relatively high change (10 percent) does not lead to any reaction. There is no doubt that this type of data should be assigned to a separate group and defined by an appropriate sensitivity indicator (VHIS). Unfortunately, the question of the remaining data is more problematic. As a solution to this problem, a quartile's function is used to divide this data into four equal groups. In this way, the five values [in percent] are obtained 0.0054 (zero quartile, minimum); 0.29 (first quartile 1); 0.98 (second quartile); 4.37 (third quartile) and 10.12 (fourth quartile, maximum) which include an equal number of data points. These values are slightly smoothed and used to form the appropriate ranges for the remaining sensitivity indicators. In this way, the five groups of data and the sensitivity indicators are obtained; then these values are assigned to DQIs and Quality Classes. Each of the group includes data with a strictly defined share in the environmental impact (calculated as the mean and median on the life cycle stage and the entire life cycle level, separately) as is shown in **Table 1**.

As one can see in the table above, the data with the lowest share in the final LCA results are assigned to the worst quality class (E) and those with the highest contribution to the best class (A). For example, data inputs which are assigned to class A show the highest degree of change in the sensitivity analysis (between 4.5 and 10.5 percent) and that is why they are assigned to the highest sensitivity indicator HS. This means that this group of data is the most important and has the highest share in the final LCA results (mean 76.4 percent and median 99.9). For these reasons the data should have the highest quality and values of DQI (5). It is valid for other cases as well. The disproportion between values of mean and median, as shown in Table 1, requires some additional explanation. It indicates substantial heterogeneity resulting from the fact that one of the data inputs, which has a (very) high share in the environmental impact on the level of life cycle stage, simultaneously has a very low share on the level of the whole life cycle. Summing up, the sensitivity analysis is used to determine minimal levels of quality of input data. It is assumed that data with high and very high insensitivity are so unimportant that even low quality levels would be acceptable. Of course, the best option would be if all data had the highest quality. In practise however, it is very difficult to carry out.

3 Data Quality Distance (DQD) and Pedigree Matrix

The next question to answer is: what is the 'real' level of quality of data? In order to solve this problem the Weidema Pedigree Matrix is used [4,5,16]. A slight modification is made by introducing some values of DQIs for each small cell of the Matrix (from 1 to 0.2), as is presented in **Table 2**. Some default requirements (ideal conditions) are assumed and called Data Quality Goals (DQGs). The highest values of DQIs - are assigned to the cells of DQGs. In the other cells, the values are respectively lower. The proper cells are chosen during the horizontal calculation and the difference between DQGs and the selected DQIs is calculated. In such a way, the values of parameter called Data Quality Distance (DQD) are obtained for each criterion. Finally, the values are automatically summed

Table 1: Relationship between the share of the input data in the environmental impact (in percent) and the results of the sensitivity analysis, sensitivity indicators, data quality indicators and quality classes

Share in the final environmental impact				Sensitivity Analysis			
On the level of life cycle's stages		On the level of the whole life cycle		Range of the changes [%]	Criterion [SI]	Indicator [DQI]	Quality Classes
Mean [%]	Median [%]	Mean [%]	Median [%]				
76.400	99.900	53.0000	98.9000	(4.5; 10.5>	HS	5	A
29.400	20.700	0.22000	0.13000	(1.0; 4.5>	LS	4	B
2.1500	0.3900	0.01600	0.00700	(0.3; 1.0>	LIS	3	C
0.5700	0.0430	0.00270	0.00160	(0; 0.3>	HIS	2	D
0.0082	0.0015	0.00012	0.00002	(0; 0>	VHIS	1	E

Table 2: Weidema pedigree matrix (1996) with a slight modification (an exemplary calculation)

Criterion	DQG	5	DQI	4	DQI	3	DQI	2	DQI	1	DQD
Reliability of source	Verified data based on measurements	1	Verified data based partly on assumptions or non-verified data based on measurements	0.8	Unverified data based partly on assumptions	0.6	Qualified estimation	0.4	Unqualified estimation	0.2	**0.4**
Completeness	Representativ e data from an adequate sample of sites over an adequate period	1	Representativ e data from a smaller number of sites over an adequate period	0.8	Representativ e data from an adequate number of sites, but over a shorter period	0.6	Representativ e data from a small number of sites over a shorter period or inadequate data from adequate number of sites	0.4	Unknown or incomplete data from a small number of sites	0.2	**0.6**
Temporal correlation	< 3 years difference	1	< 6 years difference	0.8	< 10 years difference	0.6	< 15 years difference	0.4	Unknown or > 15 years	0.2	**0.2**
Geographical correlation	Data from an adequate area	1	Average data from a larger area	0.8	Data from an area with a similar production structure	0.6	Data from an area with a slightly similar structure of production	0.4	Unknown or different data	0.2	**0.2**
Technological correlation	Data from processes studied and company-specific	1	Data from processes studied for different companies	0.8	Data from processes studied with different technology	0.6	Data from related processes and materials, the same technology	0.4	Data from related processes and materials, different technology	0.2	**0.8**
										Total DQD	**2.2**
										Quality CLASS	**C**

vertically and assigned to the appropriate quality class. All the calculations are made by the simple programme (Microsoft Visual Basic 6.0). The higher the value of DQD (differences between the requirements and the 'real' conditions), the lower the quality of data and quality class are. In this way, DQD with DQIs and Quality Classes are linked.

4 Sensitivity Analysis and Pedigree Matrix Together

On the one hand, thanks to the sensitivity analysis and SIs, the minimal levels of data quality (expressed in DQIs and Quality Classes) are obtained. On the other hand, thanks to the Pedigree Matrix and DQDs, the 'real' levels of quality for the same data (expressed also in DQIs and Quality Classes) are gained. There is a difference in the quantity and kind of the analysed data. The sensitivity analysis is carried out only for first-order data, while the analysis with Pedigree Matrix is made for first-, second- and third-order data. The main reason to do this was, that the LCA results are usually calculated not only for the main data (first-order), but for the whole second- and third-order processes connected with them. Each input data reflects some group of processes and, in this way, it contributes to the final results. That is why changes in the results of the sensitivity analysis do not only relate to the changes in first-order data, but in fact to changes in the whole processes. It does not apply to the case of the analysis based on the Pedigree Matrix. Here, every data can have an individual level of quality. Relationships between the results of the sensitivity analysis and the Pedigree Matrix are shown in **Table 3**.

The main advantage of the approach presented in this paper is that one can check whether the 'real' quality of data input is not lower than its minimal level. If yes, it is advised to change the data. And what's even more important, there is a common basis to make such a comparison between the analysis because both are expressed in the same units (DQIs and Quality Classes).

Table 3: Sensitivity analysis and pedigree matrix in the data quality assessment (both analysis are connected by quality classes and DQIs)

Sensitivity Analysis				**Pedigree Matrix**			
Range [%]	**Criterion [SI]**	**Indicator [DQI]**	**Quality classes**	**Quality classes**	**Indicator [DQI]**	**Criterion [DQD]**	**Range**
(4.5; 10.5>	HS	5	A	A	5	DQD	<0; 0.8>
(1.0; 4.5>	LS	4	B	B	4	DQD	(0.8; 1.6>
(0.3; 1.0>	LIS	3	C	C	3	DQD	(1.6; 2.4>
(0; 0.3>	HIS	2	D	D	2	DQD	(2.4; 3.2>
(0; 0>	VHIS	1	E	E	1	DQD	(3.2; 4.0>

5 Uncertainty Analysis

In the next phase, the uncertainty analysis is developed. As mentioned earlier, various probabilistic approaches have been developed in the past. Here, let's use the link between the DQIs obtained from the Pedigree Matrix and beta probability distributions [11]. There is a close relationship between quality of data expressed in DQIs and beta probability distribution parameters. There are several reasons why to choose this approach together with beta probabilistic distributions in the assessment:

1 the uncertainty analysis can be carried out for very poor data, even data points (without any standard statistical information like: mean, standard deviation, etc). The approach is based on four values: two shape parameters (minimum and maximum) and two endpoints,
2 there is a huge diversity of shapes of the beta probabilistic distributions, thanks to this the great modelling flexibility is possible,
3 the beta distributions are well known, described and available in common programs (Excel, Statistica).

Thanks to the information on uncertainty of every input of data can be obtained. Instead of the single data points, an almost infinite quantity of data (in the range of the determined endpoints of the beta distributions) can be obtained. At this point, the Monte Carlo technique can be used, but a special algorithm should first be constructed [17]. This algorithm has the input and output variables. The variables can be random or fixed, discrete or continuous. In this case, the algorithm with two types of input variables is constructed: continuous random variables and continuous fixed variables. A single output variable has a random character. The algorithm is as follows:

$$Y = \sum_j x_j a_j \qquad (1)$$

where 'Y' is a random output variable, that in practice makes up the result of LCA, 'x_j' means the random input variable (input data with the appropriate beta distributions) and 'a_j' is a certain impact coefficient (fixed variable) which determines the size of the impact for one unit of every input. The results of Monte Carlo simulations usually have some unknown type of the probability distribution. To determine this, two 'goodness of fit' tests are used: the Kolmogorov-Smirnov and chi-squared tests [16,17].

6 Commentary

The quality assessment in the form presented here seems to be a very useful tool. It makes it possible to check the difference in data quality between what there really is and, what there should be (as minimum). For one single input data it is possible to use a sensitivity analysis to determine SI, DQI and quality class. Next, the Pedigree Matrix can be used to check its real quality and, because both analyses are based on the same units, it is possible to compare the results. This would be essential, especially for the important data inputs. If the Pedigree Matrix shows a quality level which is too low (in a comparison to the result of sensitivity analysis), additional efforts should be made for improvement. And, what is worth emphasizing, the method allows one to do it in a very quick way. Uncertainty analysis also requires some comments. The presented approach (algorithm) has some advantages and some disadvantages. The establishing of the relationship between LCI (as input data) and the LCIA results can be regarded as the main advantage. The main weakness is the lack of universality. In this form, the algorithm can be used for calculations based on only the Ecoindicator99 method. The impact coefficients are established using this methodology and cannot be related to another one. This algorithm is constructed on the basis the most aggregated results (level of Ecoindicator). The same can be done for the lower levels of the aggregation (characterization, grouping, normalization, weighting). In these cases (especially in the characterization stage), however, the complexity of the algorithm and the results must increase considerably, often to an impractical size. This problem is particularly important on the interpretation level. This is why the Monte Carlo simulations are carried out using the algorithm constructed at the final level of the LCA results.

References

[1] Lewandowska A, Foltynowicz Z (2004): Comparative LCA of Industrial Objects. Part II – Case study for chosen industrial pumps. Int J LCA, OnlineFirst < DOI: http://dx.doi.org/10.1065/lca2004.03.152.2>
[2] Björklund AE (2002): Survey of Approaches to Improve Reliability in LCA. Int J LCA 7 (2) 64–72
[3] Fava J, Jensen AA, Lindfors L, Pomper S, Smet B de, Warren J, Vigon B (1992): Life-Cycle Assessment Data Quality: A Conceptual Framework; SETAC Workshop Report, Wintergreen, USA
[4] Weidema BP, Wesnaes MS (1996): Data Quality Management for Life Cycle Inventories – An Example of Using Data Quality Indicators. Journal of Cleaner Production 4 (3–4) 167–174
[5] Van den Berg N; Huppes G, Lindeijer E, Van der Ven B, Wrisberg N. (1999): Quality Assessment for LCA. CML Report 152. CML, Leiden University, The Netherlands
[6] Ansemes A, Lighart TN (2002): Data certification for LCA comparisons: Inventory of current status and strength and weakness analysis. TNO-Report, The Netherlands
[7] Barnthouse L, Fava J, Humphreys K, Hunt R, Laibson L, Noesen S, Owens J, Todd J, Vigon B, Weitz K, Young J (1997): Life-Cycle Impact Assessment: The State-of-the-Art. SETAC Report, Pensacola, USA
[8] Huijbregts AJM (1998): Application Uncertainty and Variability in LCA. Part I: Int J LCA 3 (5) 273-280. Part II: Int J LCA 3 (6) 343–351
[9] Huijbregts AJM, Norris G, Bretz R, Ciroth A, Maurice B, Bahr B von, Weidema B, Beaufort ASH de (2001): Framework for Modelling Data Uncertainty in Life Cycle Inventories. Int J LCA 6 (3) 127–132
[10] Norris G (2002): The Many Dimensions of Uncertainty Analysis in LCA; Athena Sustainable Materials Institute. <http://www.athenasmi.ca/papers/down_papers/UncertaintyAnalysis_in_LCA.pdf>
[11] Canter K, Kennedy D, Montgomery D, Keats J, Carlyle W (2002): Screening Stochastic Life Cycle Assessment Inventory Models. Int J LCA 7 (1) 18–26
[12] International standard ISO 14043 (2000): Environmental Management – Life Cycle Assessment – Life Cycle Interpretation. The International Standardization Organization. Geneva, Switzerland
[13] International standard ISO/TR 14049 (2000): Environmental Management – Life Cycle Assessment – Examples of application of ISO 14041 to goal and scope definition and inventory analysis. The International Standardization Organization. Geneva, Switzerland
[14] Almemark M, Bjuggren C, Granath J, Olsson J, Röttorp J, Lindfors LG(2000): Analysis and Development of the Interpretation Process in LCA. IVL Rapport, Stockholm, Sweden
[15] Heijungs R, Kleijn R (2001): Numerical approaches towards life cycle interpretation: Five examples. Int J LCA 6 (3) 141–148
[16] Weidema BP, Cappellaro F, Carlson R, Notten P, Palsson AC, Patyk A, Regalini E, Sacchetto F, Scalbi S (2001): Procedural guideline for collection, treatment and quality documentation of LCA data. EU-Project CASCADE
[17] Vose D (1996): Quantitative Risk Analysis. A Guide to Monte Carlo Simulation Modelling. John Wiley & Sons

Received: July 30th, 2003
Accepted: February 18th, 2004
OnlineFirst: March 6th, 2004